C29 File Structure Analysis of Media Files Transmitted and Received Over WhatsApp

Henry L. Risemberg, MS, Texas Department of Public Safety, Austin, TX 78752; Catalin Grigoras, PhD, National Center for Media Forensics, University of Colorado Denver, Denver, CO 80204; Jeff M. Smith, MS, The MITRE Corp, McLean, VA 22182*

Learning Overview: After attending this presentation, attendees will be aware of changes to file structure and metadata found to be associated with images and audio files that have been transmitted and received over the WhatsApp social media application compared to their original counterparts.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by providing analysts and investigators with information to help better understand the process of different uploading and downloading techniques on the WhatsApp application utilizing different devices and the effects these methods have on file structure and metadata.

WhatsApp is currently the third most popular social media network and the single most popular communication application worldwide with at least 1.6 billion active users. The application has garnered attention for its end-to-end encryption and the privacy it offers to users. There have been recently documented cases of criminal and terrorist organizations using WhatsApp to communicate and share files securely. The ubiquitous nature of WhatsApp in today's society, along with its use by those with nefarious criminal intentions, highlights the relevance and importance of the findings in this study.

Key identifying features of media file structures that can be attributed to files that are transmitted and received over WhatsApp from different types of devices will be analyzed. Law enforcement and other investigative agencies can use this information to help determine the source of image and audio files acquired during the course of forensic investigations.

This presentation proposes that as an image or audio file undergoes the process of being transmitted from one user to another over the WhatsApp application, that file is imparted with metadata and traces of compression unique to that process that can later be detected and identified. A dataset of hundreds of image and audio files is created by manually transmitting and then downloading a set of original images and audio files utilizing many possible transmission methods and devices. This set of files is compared to the original recorded files, as well as to each other, to identify unique file structure characteristics. Commonalities between all transmitted files is also discussed. File characteristics such as naming convention, metadata, quantization tables, and image and audio compression are examined.

Based on the results of analysis, three different image compression schemes were detected and identified as being applied to image files transmitted over WhatsApp. These compression schemes are shown to be applied to images depending on which of three broad categories of devices are used to send the image files. Analysis of hex data associated with the transmitted audio files was conducted, and a model of the audio compression applied by WhatsApp was detected and configured for a lossy audio compression database. The information presented here can be added to a growing volume of analysis conducted on media files transmitted and received over different social media platforms.

Digital, Multimedia, WhatsApp

*Presenting Author